Title: A hybrid approach for learning concept hierarchy from Malay text using artificial immune network

Author/Authors: Mohd. Zakree Ahmad Nazri, Siti Mariyam Shamsuddin, Azuraliza Abu Bakar, Salwani Abdullah

Abstract: A concept hierarchy is an integral part of an ontology but it is expensive and time consuming to build. Motivated by this, many unsupervised learning methods have been proposed to (semi-) automatically develop a concept hierarchy. A significant work is the Guided Agglomerative Hierarchical Clustering (GAHC) which relies on linguistic patterns (i.e., hypernyms) to guide the clustering process. However, GAHC still relies on contextual features to build the concept hierarchy, thus data sparsity still remains an issue in GAHC. Artificial Immune Systems are known for robustness, noise tolerance and adaptability. Thus, an extension to the GAHC is proposed by hybridizing it with Artificial Immune Network (aiNet) which we call Guided Clustering and aiNet for Learning Concept Hierarchy (GCAINY). In this paper, we have tested GCAINY using two parameter settings. The first parameter setting is obtained from the literature as a baseline parameter setting and second is by automatic parameter tuning using Particle Swarm Optimization (PSO). The effectiveness of the GCAINY is evaluated on three data sets. For further validations, a comparison between GCAINY and GAHC has been conducted and with statistical tests showing that GCAINY increases the quality of the induced concept hierarchy. The results reveal that the parameters value found by using PSO significantly produce better concept hierarchy than the vanilla parameter. Thus it can be concluded that the proposed approach has greater ability to be used in the field of ontology learning.